Biostatistics 602 - Statistical Inference
Lecture 18
Hypothesis Testing

Hyun Min Kang

March 21th, 2013

Last Lecture

- What does mean that a statistic is asymptotically normal?
- What kind of tools are useful for obtaining parameters for asymptotic normal distributions?
- How can you evaluate whether a consistent estimator is better than another consistent estimator?
- What is the Asymptotic Relative Efficiency?
- What does mean that a statistic is asymptotically efficient?
- Is an MLE asymptotically efficient?

Asymptotic Normality

Definition: Asymptotic Normality

A statistic (or an estimator) $W_n(\mathbf{X})$ is *asymptotically normal* if

$$\sqrt{n}(W_n - \tau(\theta)) \xrightarrow{d} \mathcal{N}(0, \nu(\theta))$$

for all θ
where $\xrightarrow{d}$ stands for "converge in distribution"

- $\tau(\theta)$: "asymptotic mean"
- $\nu(\theta)$: "asymptotic variance"

We denote $W_n \sim \mathcal{AN}\left(\tau(\theta), \frac{\nu(\theta)}{n}\right)$.

Central Limit Theorem

Central Limit Theorem

Assume $X_i \overset{\text{i.i.d.}}{\sim} f(x|\theta)$ with finite mean $\mu(\theta)$ and variance $\sigma^2(\theta)$.

$$\overline{X} \sim \mathcal{AN}\left(\mu(\theta), \frac{\sigma^2(\theta)}{n}\right)$$

$$\Leftrightarrow \sqrt{n}\left(\overline{X} - \mu(\theta)\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\theta))$$

Theorem 5.5.17 - Slutsky's Theorem

If $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{P} a$, where a is a constant,

1. $Y_n \cdot X_n \xrightarrow{d} aX$
2. $X_n + Y_n \xrightarrow{d} X + a$

Delta Method

Theorem 5.5.24 - Delta Method

Assume $W_n \sim \mathcal{AN}\left(\theta, \frac{\nu(\theta)}{n}\right)$. If a function g satisfies $g'(\theta) \neq 0$, then

$$g(W_n) \sim \mathcal{AN}\left(g(\theta), [g'(\theta)]^2 \frac{\nu(\theta)}{n}\right)$$

Asymptotic Efficiency

Definition : Asymptotic Efficiency for iid samples

A sequence of estimators W_n is asymptotically efficient for $\tau(\theta)$ if for all $\theta \in \Omega$,

$$\begin{aligned} \sqrt{n}(W_n - \tau(\theta)) &\xrightarrow{d} \mathcal{N}\left(0, \frac{[\tau'(\theta)]^2}{I(\theta)}\right) \\ \iff W_n &\sim \mathcal{AN}\left(\tau(\theta), \frac{[\tau'(\theta)]^2}{nI(\theta)}\right) \\ I(\theta) &= E\left[\left\{\frac{\partial}{\partial\theta}\log f(X|\theta)\right\}^2 \middle| \theta\right] \\ &= -E\left[\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\middle|\theta\right] \quad \text{(if interchangeability holds)} \end{aligned}$$

Note: $\frac{[\tau'(\theta)]^2}{nI(\theta)}$ is the C-R bound for unbiased estimators of $\tau(\theta)$.

Asymptotic Efficiency of MLEs

Theorem 10.1.12

Let $X_1, \cdots, X_n$ be iid samples from $f(x|\theta)$. Let $\hat{\theta}$ denote the MLE of θ. Under same regularity conditions, $\hat{\theta}$ is consistent and asymptotically normal for θ, i.e.

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I(\theta)}\right) \text{ for every } \theta \in \Omega$$

And if $\tau(\theta)$ is continuous and differentiable in θ, then

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta) &\xrightarrow{d} \mathcal{N}\left(0, \frac{[\tau'(\theta)]}{I(\theta)}\right) \\ \implies \tau(\hat{\theta}) &\sim \mathcal{AN}\left(\tau(\theta), \frac{[\tau'(\theta)]^2}{nI(\theta)}\right) \end{aligned}$$

Again, note that the asymptotic variance of $\tau(\hat{\theta})$ is Cramer-Rao lower bound for unbiased estimators of $\tau(\theta)$.

Hypothesis Testing

Definition

A *hypothesis* is a statement about a population parameter

Two complementary statements about θ

- Null hypothesis : $H_0 : \theta \in \Omega_0$
- Alternative hypothesis : $H_1 : \theta \in \Omega_0^c$

$\theta \in \Omega = \Omega \cup \Omega^c$.

Simple and composite hypothesis

Simple hypothesis

Both H_0 and H_1 consist of only one parameter value.

- $H_0 : \theta = \theta_0 \in \Omega_0$
- $H_1 : \theta = \theta_1 \in \Omega_0^c$

Composite hypothesis

One or both of H_0 and H_1 consist more than one parameter values.

- One-sided hypothesis: $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$.
- One-sided hypothesis: $H_0 : \theta \geq \theta_0$ vs $H_1 : \theta < \theta_0$.
- Two-sided hypothesis: $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$.

An Example of Hypothesis

$$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, 1)$$

Let X_i is the change in blood pressure after a treatment.

$$\begin{aligned} H_0 &: \theta = 0 \quad \text{(no effect)} \\ H_1 &: \theta \neq 0 \quad \text{(some effect)} \end{aligned}$$

Two-sided composite hypothesis.

Another Example of Hypothesis

- Let θ denotes the proportion of defective items from a machine.
- One may want the proportion to be less than a specified maximum acceptable proportion θ_0.
- We want to test whether the products produced by the machine is acceptable.

$$\begin{aligned} H_0 &: \theta \leq \theta_0 \quad \text{(acceptable)} \\ H_1 &: \theta > \theta_0 \quad \text{(unacceptable)} \end{aligned}$$

Hypothesis Testing Procedure

A hypothesis testing procedure is a rule that specifies:

1. For which sample points H_0 is accepted as true (the subset of the sample space for which H_0 is accepted is called the acceptable region).
2. For which sample points H_0 is rejected and H_1 is accepted as true (the subset of sample space for which H_0 is rejected is called the rejection region or critical region).

Rejection region (R) on a hypothesis is usually defined through a test statistic $W(\mathbf{X})$. For example,

$$\begin{aligned} R_1 &= \{\mathbf{x} : W(\mathbf{x}) > c, \mathbf{x} \in \mathcal{X}\} \\ R_2 &= \{\mathbf{x} : W(\mathbf{x}) \leq c, \mathbf{x} \in \mathcal{X}\} \end{aligned}$$

Example of hypothesis testing

$X_1, X_2, X_3 \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$. Consider hypothesis tests

$$\begin{aligned} H_0 &: p \leq 0.5 \\ H_1 &: p > 0.5 \end{aligned}$$

- Test 1 : Reject H_0 if $\mathbf{x} \in \{(1,1,1)\}$
  $\Longleftrightarrow$ rejection region $= \{(1,1,1)\}$
  $\Longleftrightarrow$ rejection region $= \{\mathbf{x} : \sum x_i > 2\}$
- Test 2 : Reject H_0 if $\mathbf{x} \in \{(1,1,0),(1,0,1),(0,1,1),(1,1,1)\}$
  $\Longleftrightarrow$ rejection region $= \{(1,1,0),(1,0,1),(0,1,1),(1,1,1)\}$
  $\Longleftrightarrow$ rejection region $= \{\mathbf{x} : \sum x_i > 1\}$

Example

Let $X_1, \cdots, X_n$ be changes in blood pressure after a treatment.

$$\begin{aligned} H_0 &: \theta = 0 \\ H_1 &: \theta \neq 0 \end{aligned}$$

An example rejection region $R = \left\{\mathbf{x} : \frac{\bar{x}}{s_{\mathbf{X}}/\sqrt{n}} > 3\right\}$.

Truth \ Decision	Accept H_0	Reject H_0
H_0	Correct Decision	Type I error
H_1	Type II error	Correct Decision

Type I and Type II error

Type I error

If $\theta \in \Omega_0$ (if the null hypothesis is true), the probability of making a type I error is

$$\Pr(\mathbf{X} \in R|\theta)$$

Type II error

If $\theta \in \Omega_0^c$ (if the alternative hypothesis is true), the probability of making a type II error is

$$\Pr(\mathbf{X} \notin R|\theta) = 1 - \Pr(\mathbf{X} \in R|\theta)$$

Power function

Definition - The power function

The power function of a hypothesis test with rejection region R is the function of θ defined by

$$\beta(\theta) = \Pr(\mathbf{X} \in R|\theta) = \Pr(\text{reject } H_0|\theta)$$

If $\theta \in \Omega_0^c$ (alternative is true), the probability of rejecting H_0 is called the power of test for this particular value of θ.

- Probability of type I error $= \beta(\theta)$ if $\theta \in \Omega_0$.
- Probability of type II error $= 1 - \beta(\theta)$ if $\theta \in \Omega_0^c$.

An ideal test should have power function satisfying $\beta(\theta) = 0$ for all $\theta \in \Omega_0$, $\beta(\theta) = 1$ for all $\theta \in \Omega_0^c$, which is typically not possible in practice.

Example of power function

Problem

$X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(\theta)$ where $n = 5$.

$$H_0 : \theta \leq 0.5$$
$$H_1 : \theta > 0.5$$

Test 1 rejects H_0 if and only if all "success" are observed. i.e.

$$R = \{\mathbf{x} : \mathbf{x} = (1,1,1,1,1)\} = \{\mathbf{x} : \sum_{i=1}^{5} x_i = 5\}$$

1. Compute the power function
2. What is the maximum probability of making type I error?
3. What is the probability of making type II error if $\theta = 2/3$?

Solution for Test 1

Power function

$$\beta(\theta) = \Pr(\text{reject } H_0|\theta) = \Pr(\mathbf{X} \in R|\theta) = \Pr(\sum X_i = 5|\theta)$$

Because $\sum X_i \sim \text{Binomial}(5, \theta)$, $\beta(\theta) = \theta^5$.

Maximum type I error

When $\theta \in \Omega_0 = (0, 0.5]$, the power function $\beta(\theta)$ is Type I error.

$$\max_{\theta \in (0,0.5]} \beta(\theta) = \max_{\theta \in (0,0.5]} \theta^5 = 0.5^5 = 1/32 \approx 0.031$$

Type II error when $\theta = 2/3$

$$1 - \beta(\theta)|_{\theta=\frac{2}{3}} = 1 - \theta^5\big|_{\theta=\frac{2}{3}} = 1 - (2/3)^5 = 211/243 \approx 0.868$$

Another Example

Problem

$X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(\theta)$ where $n = 5$.

$$H_0 : \theta \leq 0.5$$
$$H_1 : \theta > 0.5$$

Test 2 rejects H_0 if and only if 3 or more "success" are observed. i.e.

$$R = \{\mathbf{x} : \sum_{i=1}^{5} x_i \geq 3\}$$

1. Compute the power function
2. What is the maximum probability of making type I error?
3. What is the probability of making type II error if $\theta = 2/3$?

Solution for Test 2

Power function

$$\beta(\theta) = \Pr(\sum X_i \geq 3|\theta) = \binom{5}{3}\theta^3(1-\theta)^2 + \binom{5}{4}\theta^4(1-\theta) + \binom{5}{5}\theta^5 = \theta^3(6\theta^2 - 15\theta + 10)$$

Maximum type I error

We need to find the maximum of $\beta(\theta)$ for $\theta \in \Omega_0 = (0, 0.5]$

$$\beta'(\theta) = 30\theta^2(\theta - 1)^2 > 0$$

$\beta(\theta)$ is increasing in $\theta \in (0, 1)$. Maximum type I error is $\beta(0.5) = 0.5$

Type II error when $\theta = 2/3$

$$1 - \beta(\theta)|_{\theta=\frac{2}{3}} = 1 - \theta^3(6\theta^2 - 15\theta + 10)\big|_{\theta=\frac{2}{3}} \approx 0.21$$

Sizes and Levels of Tests

Size α test

A test with power function $\beta(\theta)$ is a size α test if

$$\sup_{\theta \in \Omega_0} \beta(\theta) = \alpha$$

In other words, the maximum probability of making a type I error is α.

Level α test

A test with power function $\beta(\theta)$ is a level α test if

$$\sup_{\theta \in \Omega_0} \beta(\theta) \leq \alpha$$

In other words, the maximum probability of making a type I error is equal or less than α.

Any size α test is also a level α test

Revisiting Previous Examples

Test 1

$$\sup_{\theta \in \Omega_0} \beta(\theta) = \sup_{\theta \in \Omega_0} \theta^5 = 0.5^5 = 0.03125$$

The size is 0.03125, and this is a level 0.05 test, or a level 0.1 test, but not a level 0.01 test.

Test 2

$$\sup_{\theta \in \Omega_0} \beta(\theta) = 0.5$$

The size is 0.5

Constructing a good test

1. Construct all the level α test.
2. Within this level of tests, we search for the test with Type II error probability as small as possible; equivalently, we want the test with the largest power if $\theta \in \Omega_0^c$.

Review on standard normal and t distribution

Quantile of standard normal distribution

Let $Z \sim \mathcal{N}(0,1)$ with pdf $f_Z(z)$ and cdf $F_Z(z)$. The α-th quantile z_α or $(1-\alpha)$-th quantile $z_{1-\alpha}$ of the standard distribution satisfy

$$\Pr(Z \geq z_\alpha) = \alpha \quad \text{or} \quad z_\alpha = F_Z^{-1}(1-\alpha)$$
$$\Pr(Z \leq z_{1-\alpha}) = \alpha \quad \text{or} \quad z_{1-\alpha} = F_Z^{-1}(\alpha)$$
$$z_{1-\alpha} = -z_\alpha$$

Quantile of t distribution

Let $T \sim t_{n-1}$ with pdf $f_{T,n-1}(t)$ and cdf $F_{T,n-1}(t)$. The α-th quantile $t_{n-1,\alpha}$ or $(1-\alpha)$-th quantile $t_{n-1,1-\alpha}$ of the standard distribution satisfy

$$\Pr(T \geq t_{n-1,\alpha}) = \alpha \quad \text{or} \quad t_{n-1\alpha} = F_{T,n-1}^{-1}(1-\alpha)$$
$$\Pr(T \leq t_{n-1,1-\alpha}) = \alpha \quad \text{or} \quad t_{n-1,1-\alpha} = F_{T,n-1}^{-1}(\alpha)$$
$$t_{n-1,1-\alpha} = -t_{n-1,\alpha}$$

Likelihood Ratio Tests (LRT)

Definition

Let $L(\theta|\mathbf{x})$ be the likelihood function of θ. The likelihood ratio test statistic for testing $H_0 : \theta \in \Omega_0$ vs. $H_1 : \theta \in \Omega_0^c$ is

$$\lambda(\mathbf{x}) = \frac{\sup_{\theta\in\Omega_0} L(\theta|\mathbf{x})}{\sup_{\theta\in\Omega} L(\theta|\mathbf{x})} = \frac{L(\hat{\theta}_0|\mathbf{x})}{L(\hat{\theta}|\mathbf{x})}$$

where $\hat{\theta}$ is the MLE of θ over $\theta \in \Omega$, and $\hat{\theta}_0$ is the MLE of θ over $\theta \in \Omega_0$ (restricted MLE).
The *likelihood ratio test* is a test that rejects H_0 if and only if $\lambda(\mathbf{x}) \leq c$ where $0 \leq c \leq 1$.

Properties of LRT

- For example
  - If $c = 1$, null hypothesis will always be rejected.
  - If $c = 0$, null hypothesis will never be rejected.
- Difference choice of $c \in [0, 1]$ give different tests.
  - The smaller the c, the smaller type I error.
  - The larger the c, the smaller the type II error.
- Choose c such that type I error probability of LRT is bound above by α.

$$\begin{aligned}\sup_{\theta\in\Omega_0} \Pr(\lambda(\mathbf{x}) \leq c) &= \sup_{\theta\in\Omega_0} \beta(\theta) \\ &= \sup_{\theta\in\Omega_0} \Pr(\text{reject } H_0) = \alpha\end{aligned}$$

Then we get a size α test.

Example of LRT

Problem

Consider $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, \sigma^2)$ where σ^2 is known.

$$\begin{aligned}H_0 &: \theta \leq \theta_0 \\ H_1 &: \theta > \theta_0\end{aligned}$$

For the LRT test and its power function

Solution

$$\begin{aligned}L(\theta|\mathbf{x}) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i-\theta)^2}{2\sigma^2}\right] \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\frac{\sum_{i=1}^n (x_i-\theta)^2}{2\sigma^2}\right]\end{aligned}$$

We need to find MLE of θ over $\Omega = (-\infty, \infty)$ and $\Omega_0 = (-\infty, \theta_0]$.

MLE of θ over $\Omega = (-\infty, \infty)$

To maximize $L(\theta|\mathbf{x})$, we need to maximize $\exp\left[-\frac{\sum_{i=1}^n (x_i-\theta)^2}{2\sigma^2}\right]$, or equivalently to minimize $\sum_{i=1}^n (x_i - \theta)^2$.

$$\begin{aligned}\sum_{i=1}^n (x_i-\theta)^2 &= \sum_{i=1}^n (x_i^2 + \theta^2 - 2\theta x_i) \\ &= n\theta^2 - 2\theta \sum_{i=1}^n x_i + \sum_{i=1}^n x_i^2\end{aligned}$$

The equation above minimizes when $\theta = \hat{\theta} = \frac{\sum_{i=1}^n x_i}{n} = \overline{x}$.

MLE of θ over $\Omega_0 = (-\infty, \theta_0]$

- $L(\theta|\mathbf{x})$ is maximized at $\theta = \frac{\sum_{i=1}^n x_i}{n} = \overline{x}$ if $\overline{x} \leq \theta_0$.
- However, if $\overline{x} \geq \theta_0$, $\overline{x}$ does not fall into a valid range of $\hat{\theta}_0$, and $\theta \leq \theta_0$, the likelihood function will be an increasing function. Therefore $\hat{\theta}_0 = \theta_0$.

To summarize,

$$\hat{\theta}_0 = \begin{cases} \overline{X} & \text{if } \overline{X} \leq \theta_0 \\ \theta_0 & \text{if } \overline{X} > \theta_0 \end{cases}$$

Likelihood ratio test

$$\lambda(\mathbf{x}) = \frac{L(\hat{\theta}_0|\mathbf{x})}{L(\hat{\theta}|\mathbf{x})} = \begin{cases} 1 & \text{if } \overline{X} \leq \theta_0 \\ \frac{\exp\left[-\frac{\sum_{i=1}^n (x_i-\theta_0)^2}{2\sigma^2}\right]}{\exp\left[-\frac{\sum_{i=1}^n (x_i-\overline{x})^2}{2\sigma^2}\right]} & \text{if } \overline{X} > \theta_0 \end{cases}$$

$$= \begin{cases} 1 & \text{if } \overline{X} \leq \theta_0 \\ \exp\left[-\frac{n(\overline{x}-\theta_0)^2}{2\sigma^2}\right] & \text{if } \overline{X} > \theta_0 \end{cases}$$

Therefore, the likelihood test rejects the null hypothesis if and only if

$$\exp\left[-\frac{n(\overline{x}-\theta_0)^2}{2\sigma^2}\right] \leq c$$

and $\overline{x} \geq \theta_0$.

Specifying c

$$\begin{aligned} \exp\left[-\frac{n(\overline{x}-\theta_0)^2}{2\sigma^2}\right] &\leq c \\ \iff -\frac{n(\overline{x}-\theta_0)^2}{2\sigma^2} &\leq \log c \\ \iff (\overline{x}-\theta_0)^2 &\geq -\frac{2\sigma^2 \log c}{n} \\ \iff \overline{x}-\theta_0 &\geq \sqrt{-\frac{2\sigma^2 \log c}{n}} \qquad (\because \overline{x} > \theta_0) \end{aligned}$$

Specifying c (cont'd)

So, LRT rejects H_0 if and only if

$$\begin{aligned} \overline{x}-\theta_0 &\geq \sqrt{-\frac{2\sigma^2 \log c}{n}} \\ \iff \frac{\overline{x}-\theta_0}{\sigma/\sqrt{n}} &\geq \frac{\sqrt{-\frac{2\sigma^2 \log c}{n}}}{\sigma/\sqrt{n}} = c^* \end{aligned}$$

Therefore, the rejection region is

$$\left\{\mathbf{x} : \frac{\overline{x}-\theta_0}{\sigma/\sqrt{n}} \geq c^*\right\}$$

Power function

$$\beta(\theta) = \Pr(\text{reject } H_0) = \Pr\left(\frac{\overline{X}-\theta_0}{\sigma/\sqrt{n}} \geq c^*\right)$$
$$= \Pr\left(\frac{\overline{X}-\theta+\theta-\theta_0}{\sigma/\sqrt{n}} \geq c^*\right)$$
$$= \Pr\left(\frac{\overline{X}-\theta}{\sigma/\sqrt{n}} \geq \frac{\theta_0-\theta}{\sigma/\sqrt{n}} + c^*\right)$$

Since $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, \sigma^2)$, $\overline{X} \sim \mathcal{N}\left(\theta, \frac{\sigma^2}{n}\right)$. Therefore,

$$\frac{\overline{X}-\theta}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$
$$\Longrightarrow \beta(\theta) = \Pr\left(Z \geq \frac{\theta_0-\theta}{\sigma/\sqrt{n}} + c^*\right)$$

where $Z \sim \mathcal{N}(0,1)$.

Making size α LRT

To make a size α test,

$$\sup_{\theta\in\Omega_0} \beta(\theta) = \alpha$$
$$\sup_{\theta\leq\theta_0} \Pr\left(Z \geq \frac{\theta_0-\theta}{\sigma/\sqrt{n}} + c^*\right) = \alpha$$
$$\Pr(Z \geq c^*) = \alpha$$
$$c^* = z_\alpha$$

Note that $\Pr\left(Z \geq \frac{\theta_0-\theta}{\sigma/\sqrt{n}} + c^*\right)$ is maximized when θ is maximum (i.e. $\theta = \theta_0$).
Therefore, size α LRT test rejects H_0 if and only if $\frac{\overline{x}-\theta_0}{\sigma/\sqrt{n}} \geq z_\alpha$.

Summary

Today

- Hypothesis Testing
- Likelihood Ratio Test

Next Lecture

- More Hypothesis Testing